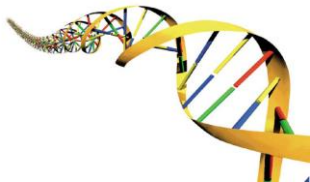
Workshop "Research as a tool for freshwater management"
Expo Aquae Venezia, September 8th, 2015

Tassonomia molecolare per razionalizzare il monitoraggio biologico

Diego Fontaneto

National Research Council, Institute of Ecosystem Study, CNR-ISE
Verbania Pallanza, Italy

d.fontaneto@ise.cnr.it



Outline

1. Species lists: surveys & biomonitoring

2. Species identification
   - Morphological taxonomy
   - DNA taxonomy
   - NGS and metagenetics

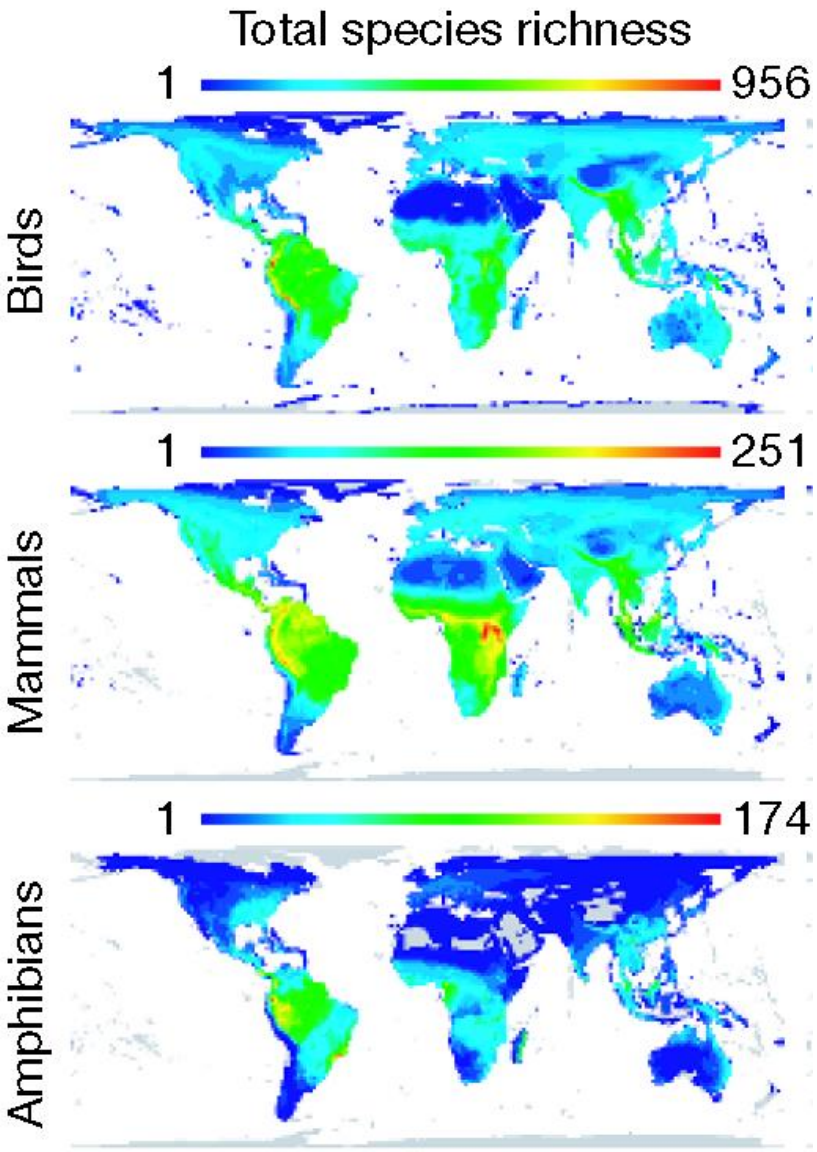
3. Inference

Outline

1. Species lists: surveys & biomonitoring

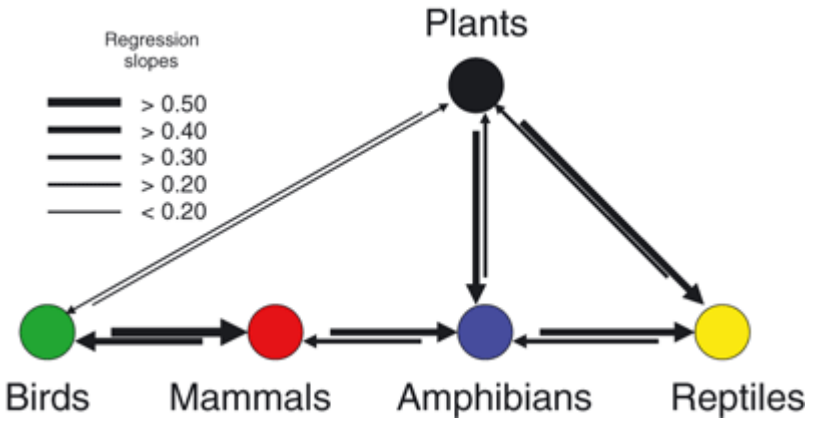
2. Species identification
   - Morphological taxonomy
   - DNA taxonomy
   - NGS and metagenetics

3. Inference



(Grenyer et al., 2006: Nature)

Global drivers of diversity



(Qian & Ricklefs, 2008: Ecology Letters)

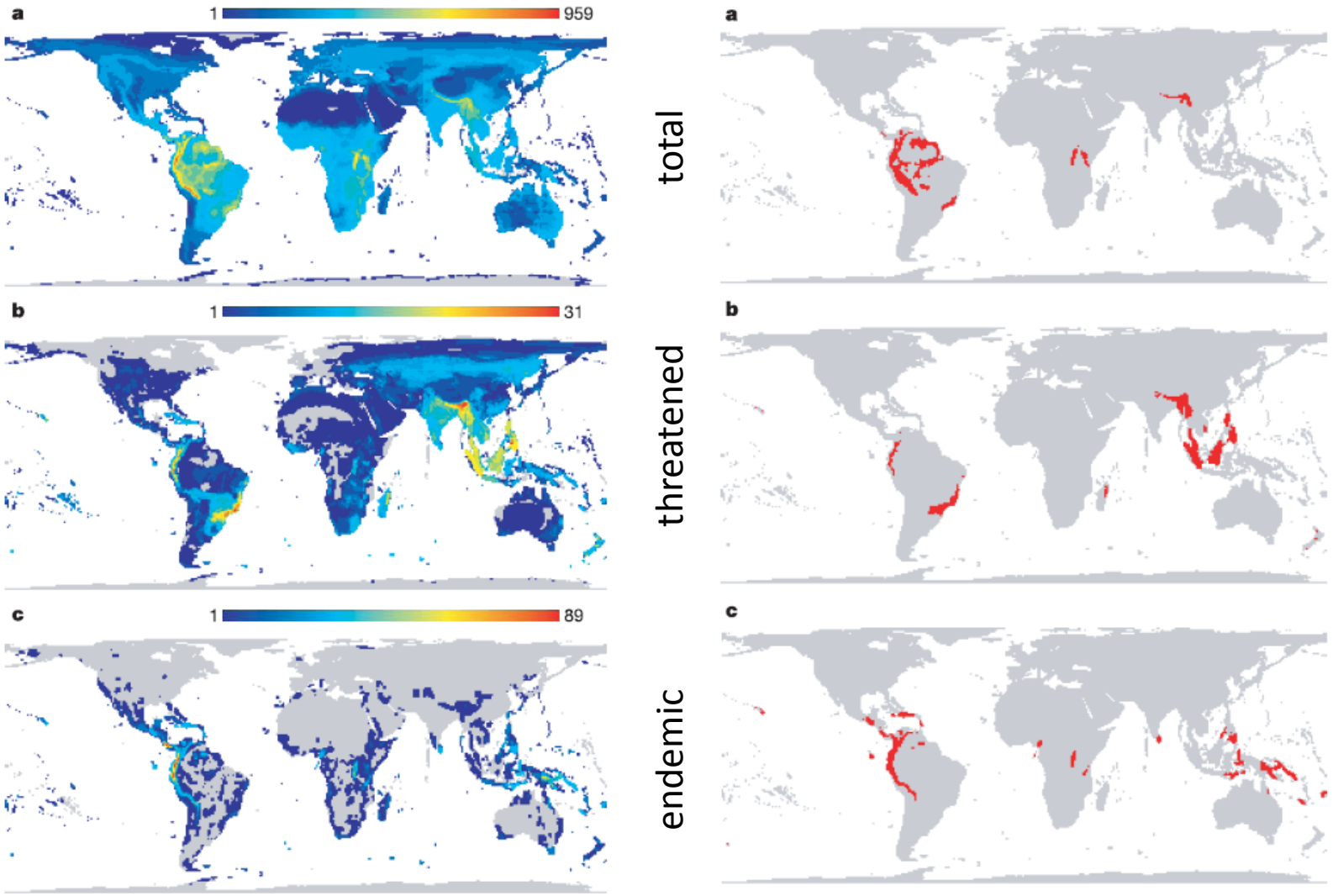
No or few data available for invertebrates

No or few data available for aquatic organisms

1. Species lists: surveys

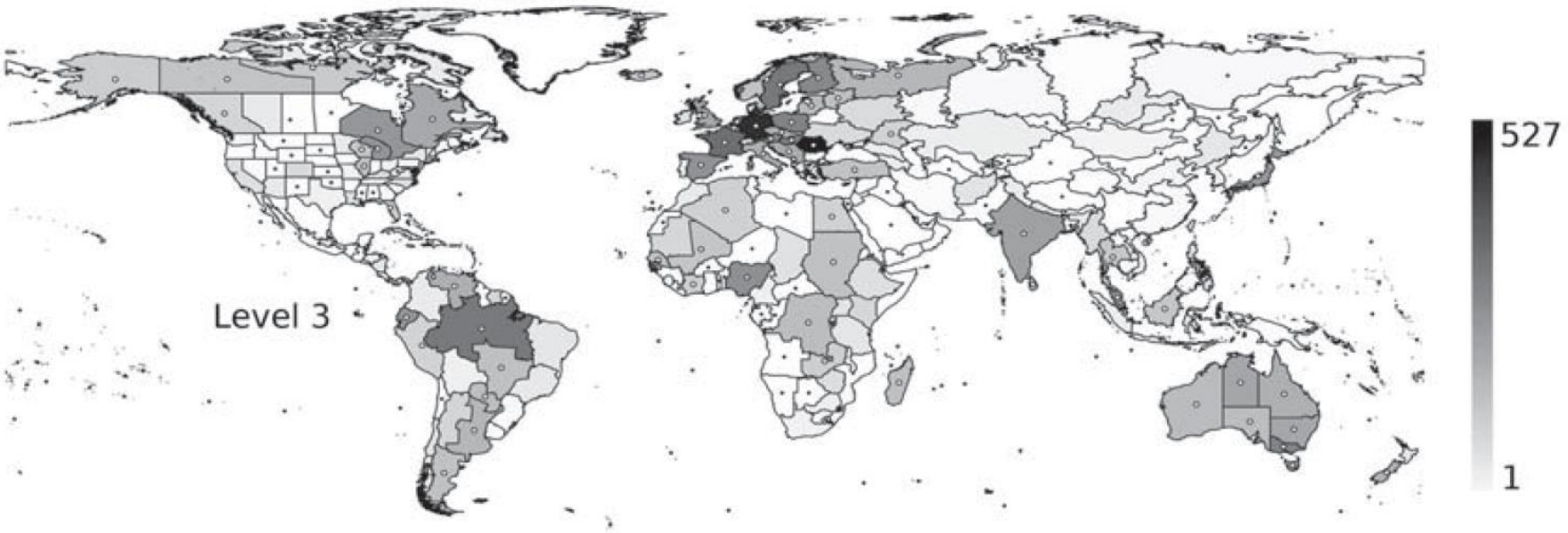
Global drivers of diversity

Bird species richness



(Orme et al., 2005: Nature)

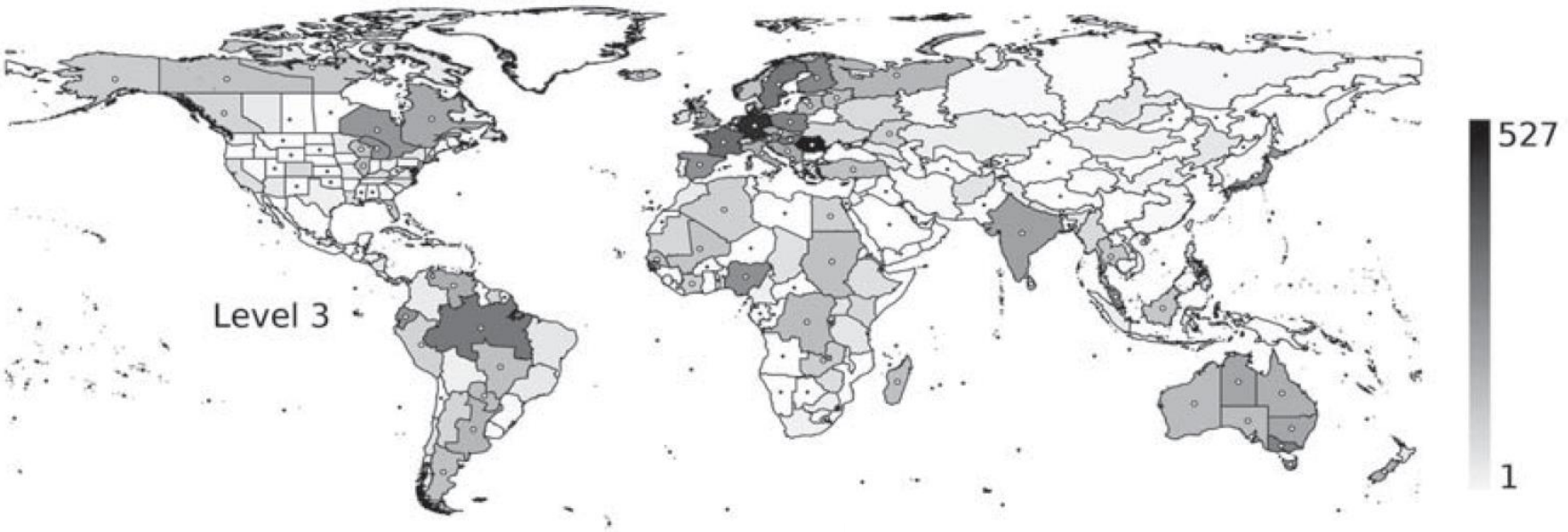
Global drivers of diversity



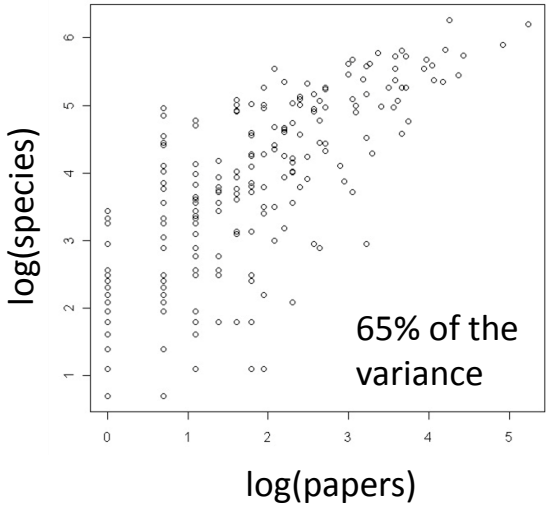
Rotifera
freshwater Monogononta
> 45,000 records
1,800 papers
from 1960 to 1993

(Fontaneto et al., 2012: Ecography)

Global drivers of diversity

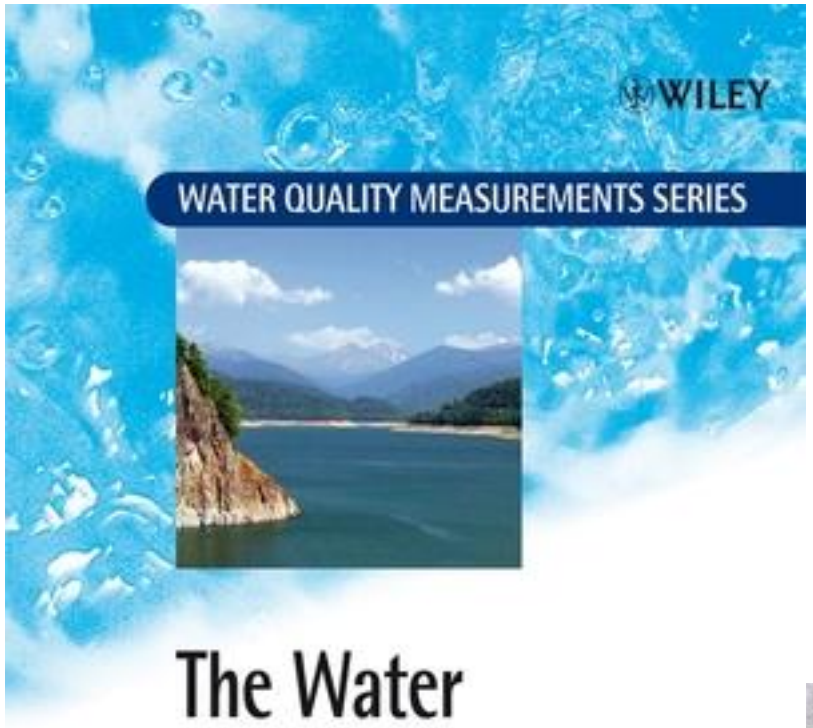


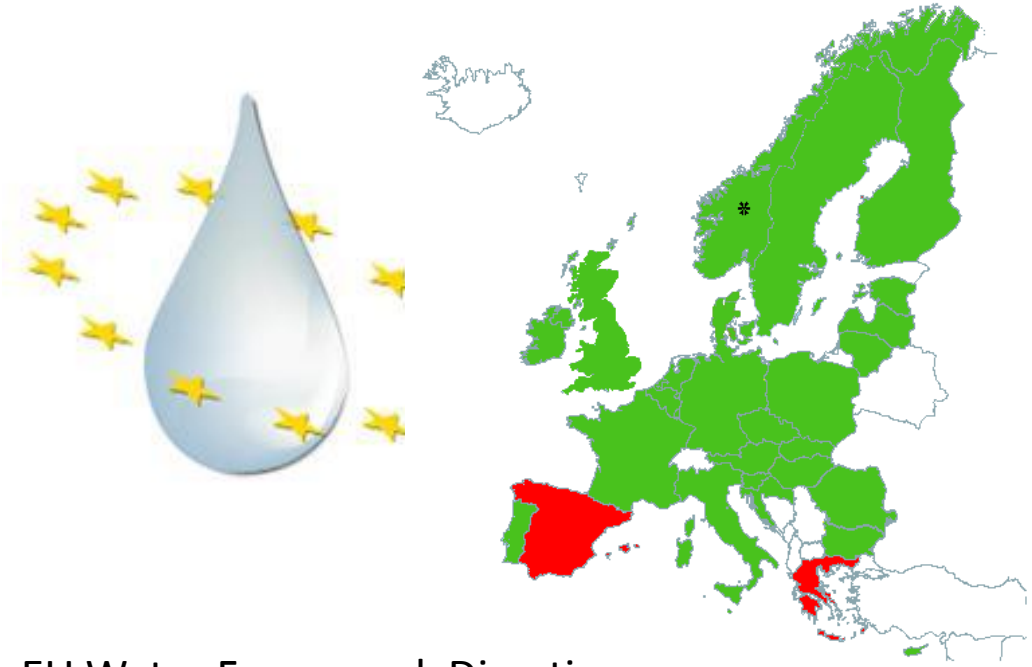
Rotifera
freshwater Monogononta
> 45,000 records
1,800 papers
from 1960 to 1993



(Fontaneto et al., 2012: Ecography)

1. Species lists: biomonitoring





EU Water Framework Directive

Outline

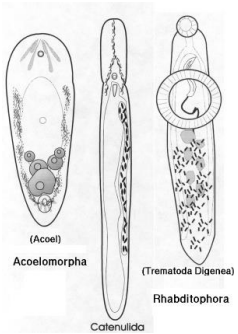
1. Species lists: surveys & biomonitoring

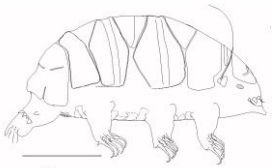
2. Species identification
   - Morphological taxonomy
   - DNA taxonomy
   - NGS and metagenetics

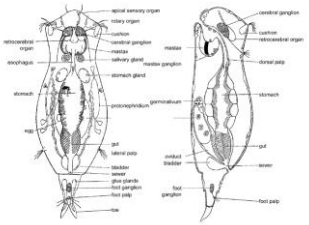
3. Inference

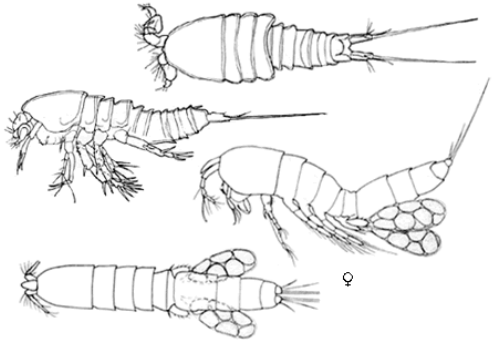
2. Species identification: morphology

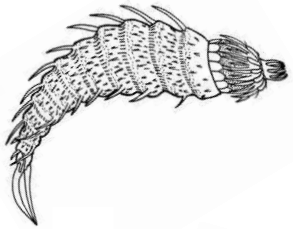
Morphological taxonomy

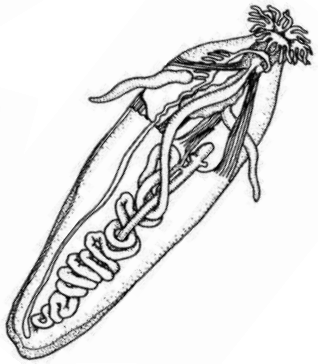


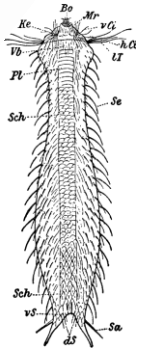


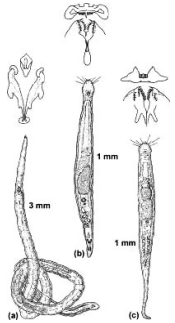


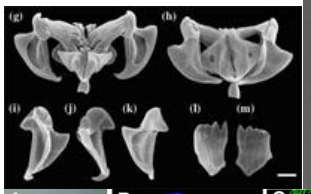


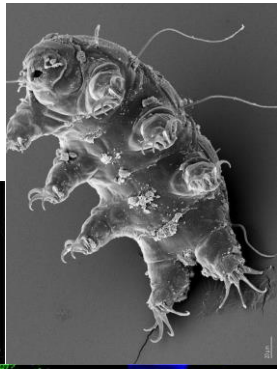


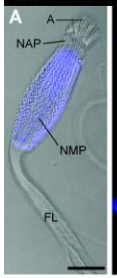


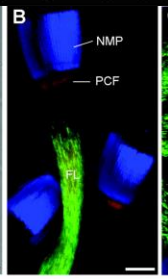


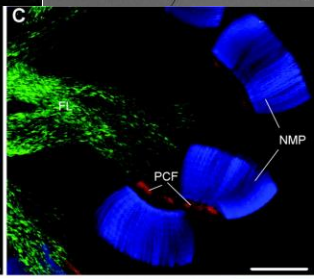


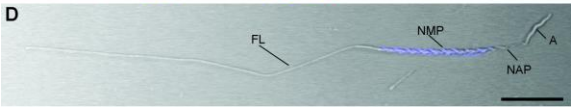


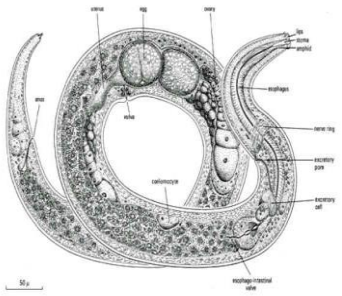


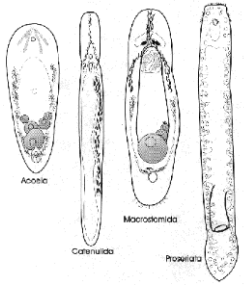


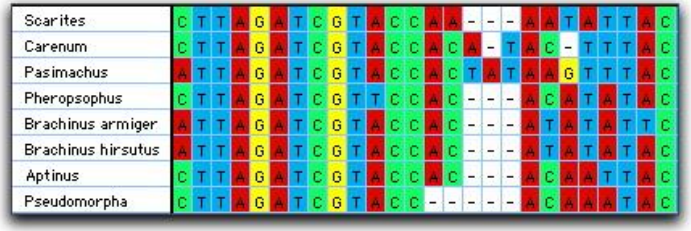












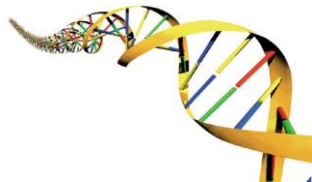
Methods in DNA taxonomy, single locus

A) Fixed thresholds, Operational Taxonomic Units (OTUs)

B) Barcoding gap, Automatic Barcode Gap Discovery (ABGD)

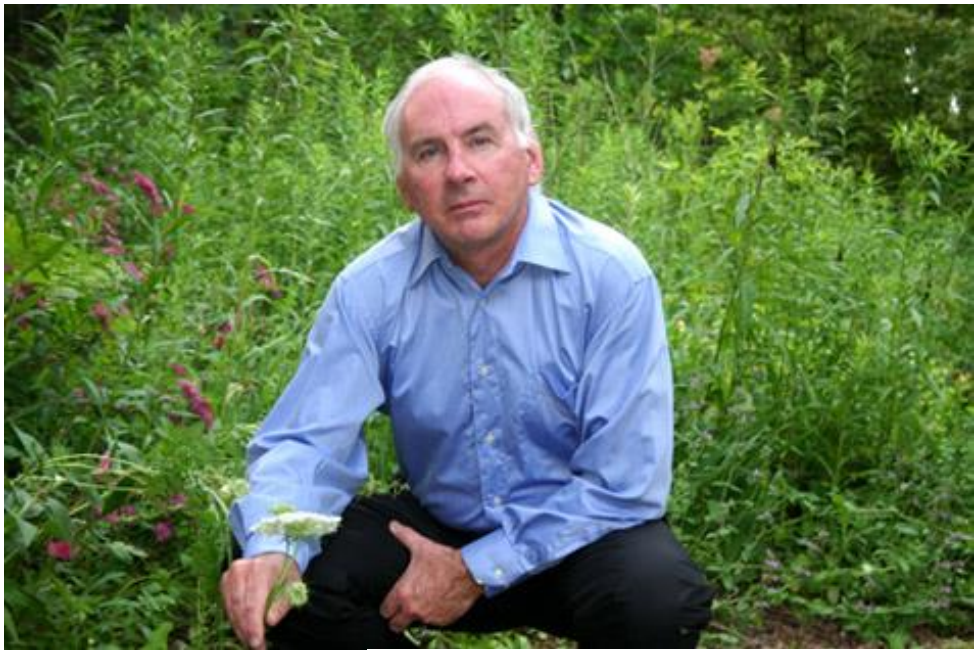
C) Tree topology, Generalized Mixed Yule Coalescent model (GMYC)

And others, such as PTP, K/theta, Haploweb, BP&P, etc…



A) Fixed thresholds, Operational Taxonomic Units (OTUs)

DNA barcoding



Paul Hebert, Guelph, Canada

THE ROYAL SOCIETY

Biological identifications through DNA barcodes

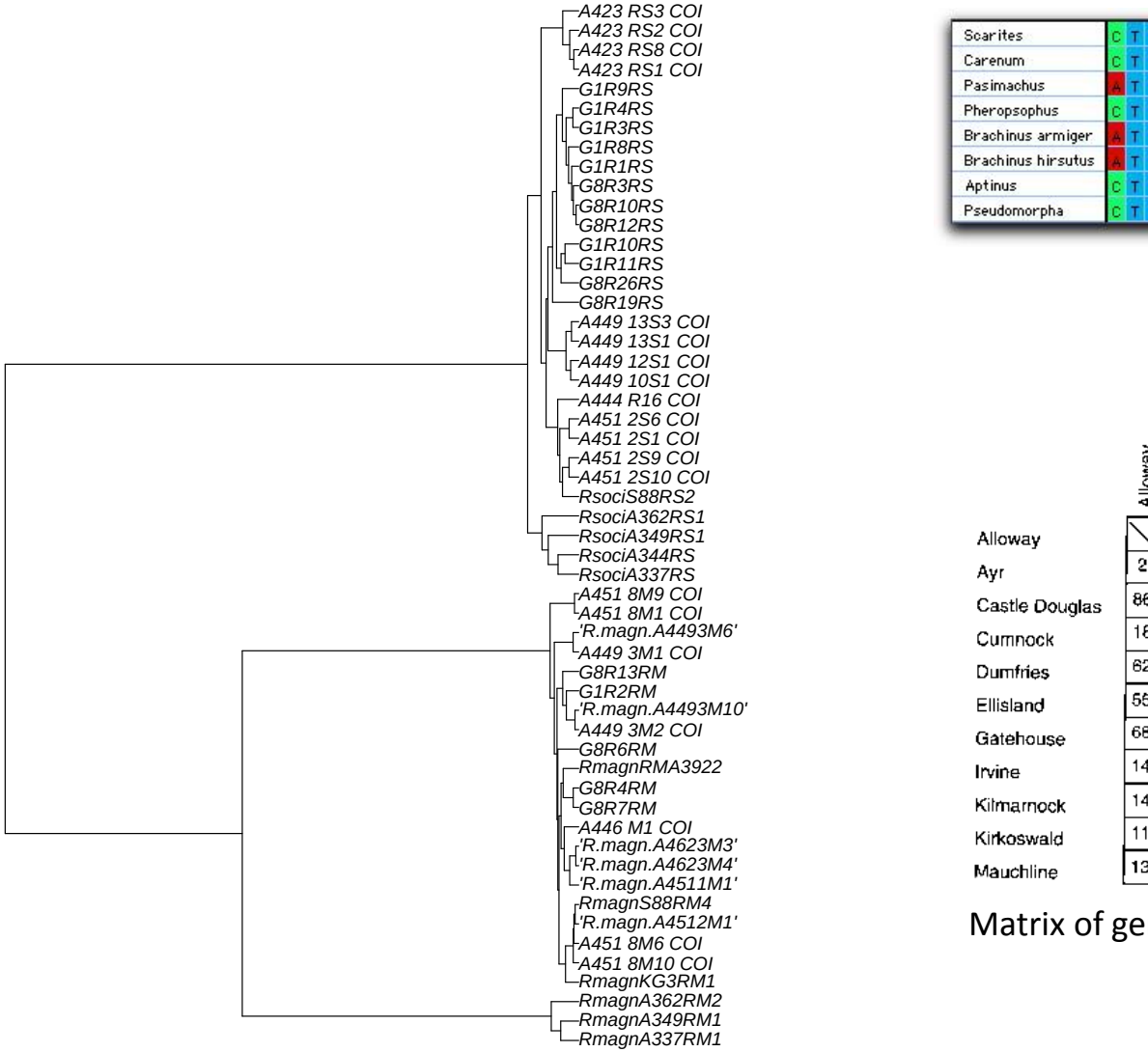
Paul D. N. Hebert*, Alina Cywinska, Shelley L. Ball and Jeremy R. deWaard

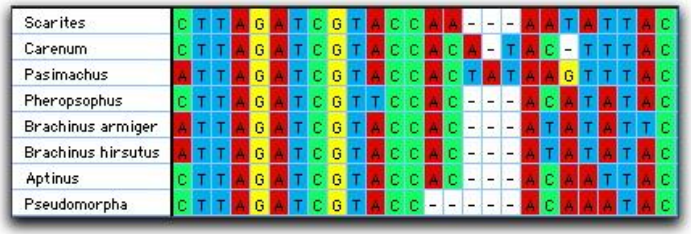
Department of Zoology, University of Guelph, Guelph, Ontario N1G 2W1, Canada

(Hebert et al., 2003: Proc. R. Soc. Lond. B)

A) Fixed thresholds, Operational Taxonomic Units (OTUs)

DNA barcoding





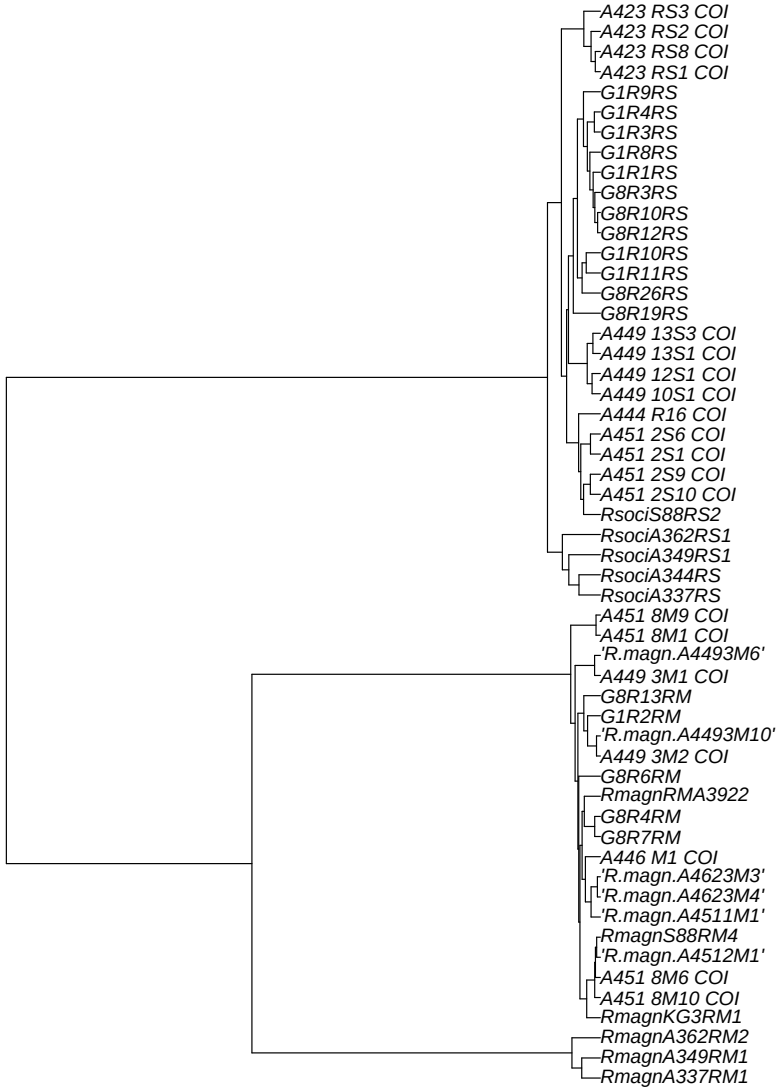
Alignment

	Alloway	Ayr	Castle Douglas	Cumnock	Dumfries	Ellisland	Gatehouse	Irvine	Kilmarnock	Kirkoswald	Mauchline
Alloway											
Ayr	2										
Castle Douglas	86	88									
Cumnock	18	16	64								
Dumfries	62	60	20	44							
Ellisland	55	53	27	37	7						
Gatehouse	68	70	18	82	38	45					
Irvine	14	12	86	28	66	59	82				
Kilmarnock	14	12	79	15	59	52	82	7			
Kirkoswald	11	13	75	29	73	66	57	25	25		
Mauchline	13	11	70	6	50	43	81	24	9	24	

Matrix of genetic distances

(Hebert et al., 2003: Proc. R. Soc. Lond. B)

A) Fixed thresholds, Operational Taxonomic Units (OTUs)



genetic distance:

3% dissimilarity
97% similarity

1% dissimilarity
99% similarity

DNA barcoding

(Hebert et al., 2003: Proc. R. Soc. Lond. B)

B) Barcoding gap, Automatic Barcode Gap Discovery (ABGD)

Automatic Barcode Gap Discovery
ABGD

Nicolas Puillandre, Paris, France

MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 1864–1877 doi: 10.1111/j.1365-294X.2011.05239.x

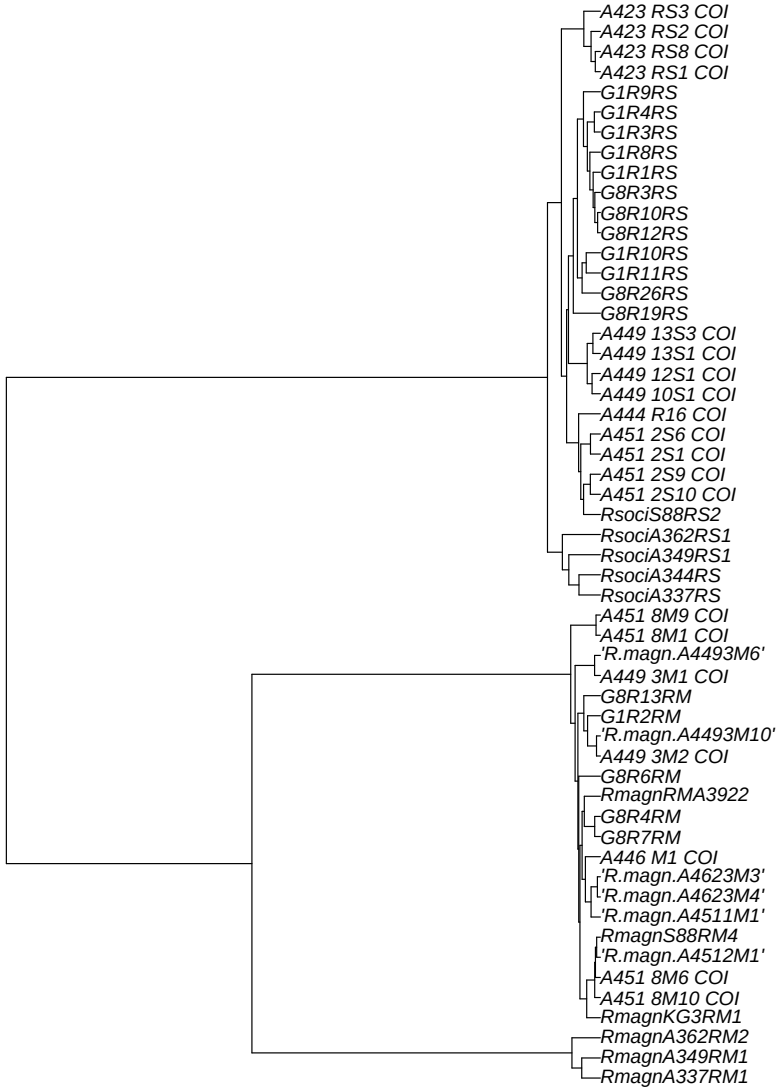
ABGD, Automatic Barcode Gap Discovery for primary species delimitation

N. PUILLANDRE,* A. LAMBERT,† S. BROUILLET‡§ and G. ACHAZ‡§

*UMR 7138, Muséum National d'Histoire Naturelle, Departement Systématique et Evolution, 43, Rue Cuvier, 75231 Paris, France, †Laboratoire de Probabilités et Modèles Aléatoires (UMR 7599), UPMC Univ Paris 06, Univ Paris Diderot, CNRS, Paris, France, ‡Systématique, Adaptation et Evolution (UMR 7138), UPMC Univ Paris 06, CNRS, MNHN, IRD, Paris, France, §Atelier de Bioinformatique, UPMC Univ Paris 06, Paris, France

(Puillandre et al., 2012: Molecular Ecology)

B) Barcoding gap, Automatic Barcode Gap Discovery (ABGD)



genetic distance:

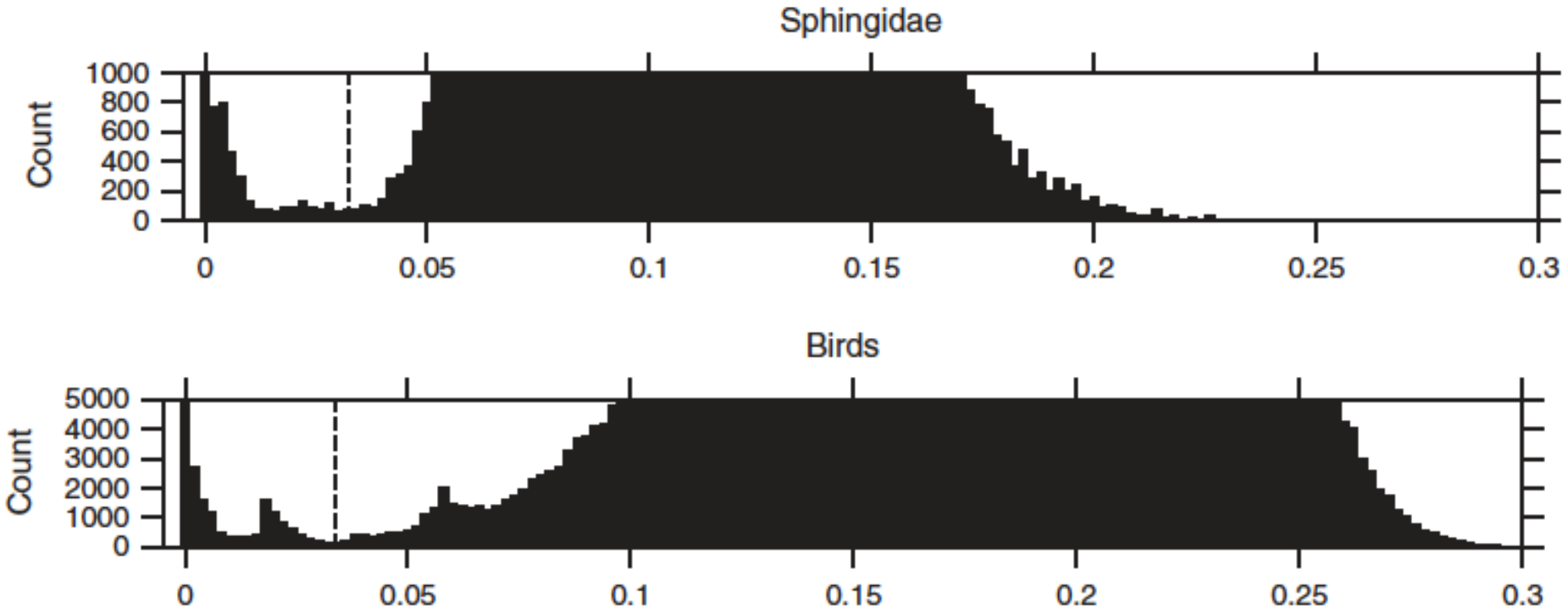
not fixed

Barcoding Gap?

	Alloway	Ayr	Castle Douglas	Cumnock	Dumfries	Ellisland	Gatehouse	Irvine	Kilmarnock	Kirkoswald	Mauchline
Alloway											
Ayr	2										
Castle Douglas	86	88									
Cumnock	18	16	64								
Dumfries	62	60	20	44							
Ellisland	55	53	27	37	7						
Gatehouse	68	70	18	82	38	45					
Irvine	14	12	86	28	66	59	82				
Kilmarnock	14	12	79	15	59	52	82	7			
Kirkoswald	11	13	75	29	73	66	57	25	25		
Mauchline	13	11	70	6	50	43	81	24	9	24	

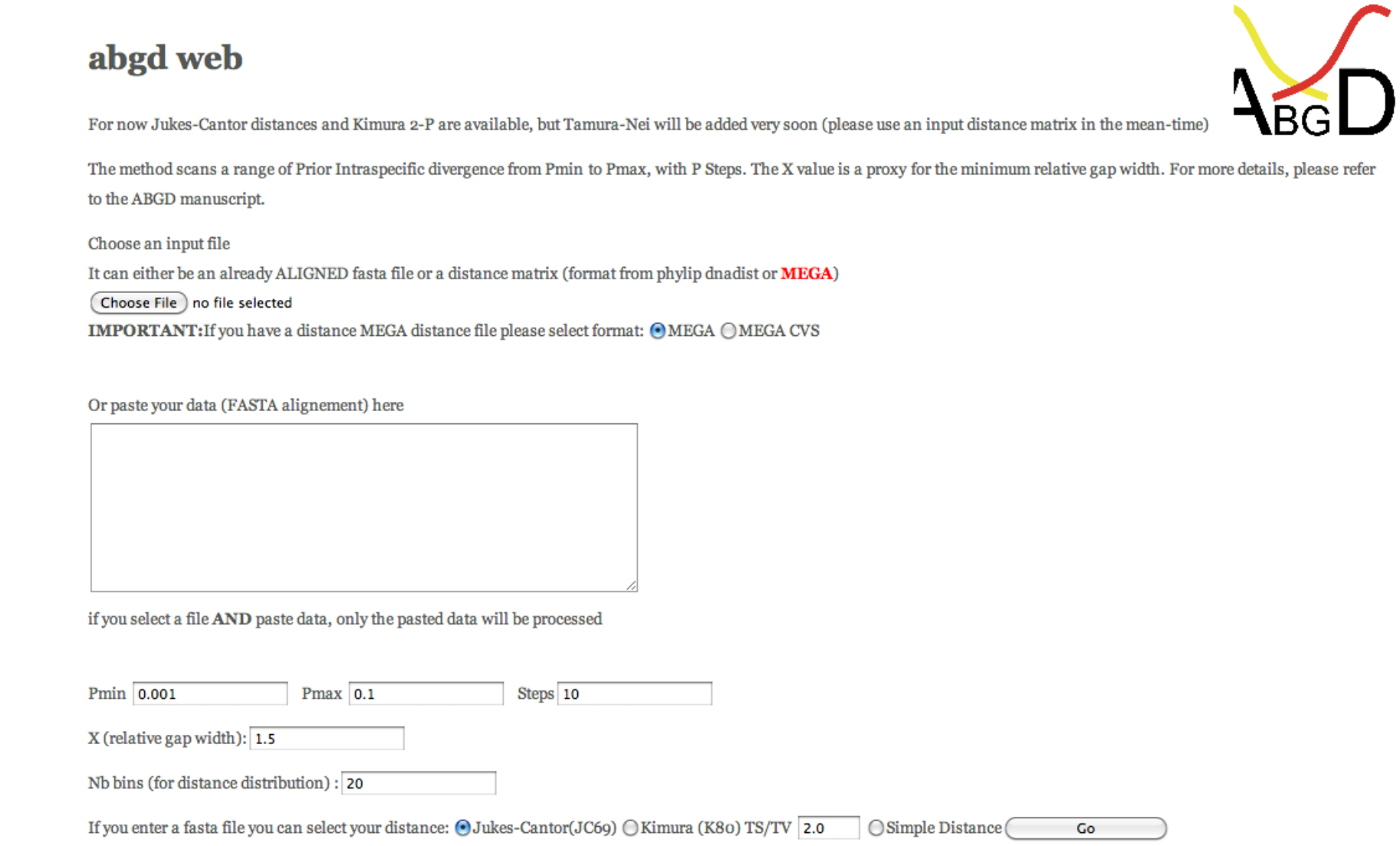
(Puillandre et al., 2012: Molecular Ecology)

B) Barcoding gap, Automatic Barcode Gap Discovery (ABGD)



(Puillandre et al., 2012: Molecular Ecology)

B) Barcoding gap, Automatic Barcode Gap Discovery (ABGD)



abgd web

For now Jukes-Cantor distances and Kimura 2-P are available, but Tamura-Nei will be added very soon (please use an input distance matrix in the mean-time)

The method scans a range of Prior Intraspecific divergence from Pmin to Pmax, with P Steps. The X value is a proxy for the minimum relative gap width. For more details, please refer to the ABGD manuscript.

Choose an input file

It can either be an already ALIGNED fasta file or a distance matrix (format from phylip dnadist or **MEGA**)

Choose File no file selected

IMPORTANT: If you have a distance MEGA distance file please select format: MEGA MEGA CVS

Or paste your data (FASTA alignement) here

if you select a file **AND** paste data, only the pasted data will be processed

Pmin 0.001 Pmax 0.1 Steps 10

X (relative gap width): 1.5

Nb bins (for distance distribution) : 20

If you enter a fasta file you can select your distance: Jukes-Cantor(JC69) Kimura (K80) TS/TV 2.0 Simple Distance Go

(Puillandre et al., 2012: Molecular Ecology)

C) Tree topology, Generalized Mixed Yule Coalescent model (GMYC)

Tim Barraclough, London, UK

Generalized Mixed Yule Coalescent model
GMYC

Syst. Biol. 62(5):707–724, 2013

DOI:10.1093/sysbio/syt033
Advance Access publication May 16, 2013

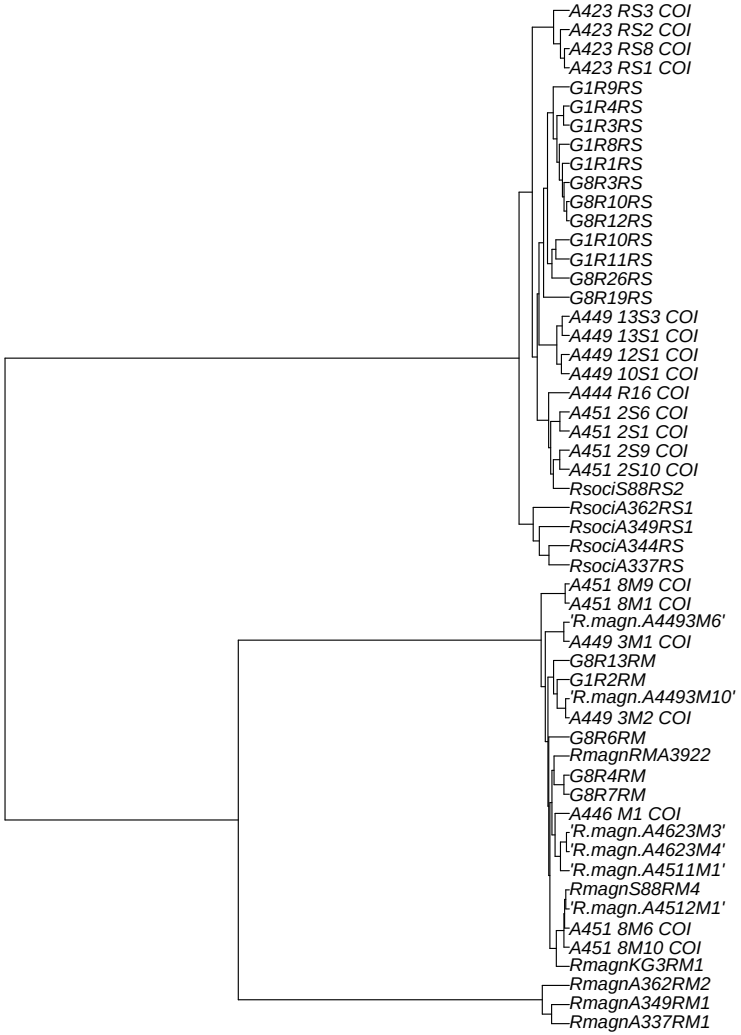
Delimiting Species Using Single-Locus Data and the Generalized Mixed Yule Coalescent Approach: A Revised Method and Evaluation on Simulated Data Sets

TOMOCHIKA FUJISAWA[1,2] AND TIMOTHY G. BARRACLOUGH[1]*

[1]*Department of Life Sciences, Imperial College London, Silwood Park Campus, Ascot, Berkshire SL5 7PY, UK; and* [2]*Department of Entomology, Natural History Museum, London SW7 5BD, UK*
**Correspondence to be sent to: Department of Life Sciences, Imperial College London, Silwood Park Campus, Ascot, Berkshire SL5 7PY, UK; E-mail: t.barraclough@imperial.ac.uk.*

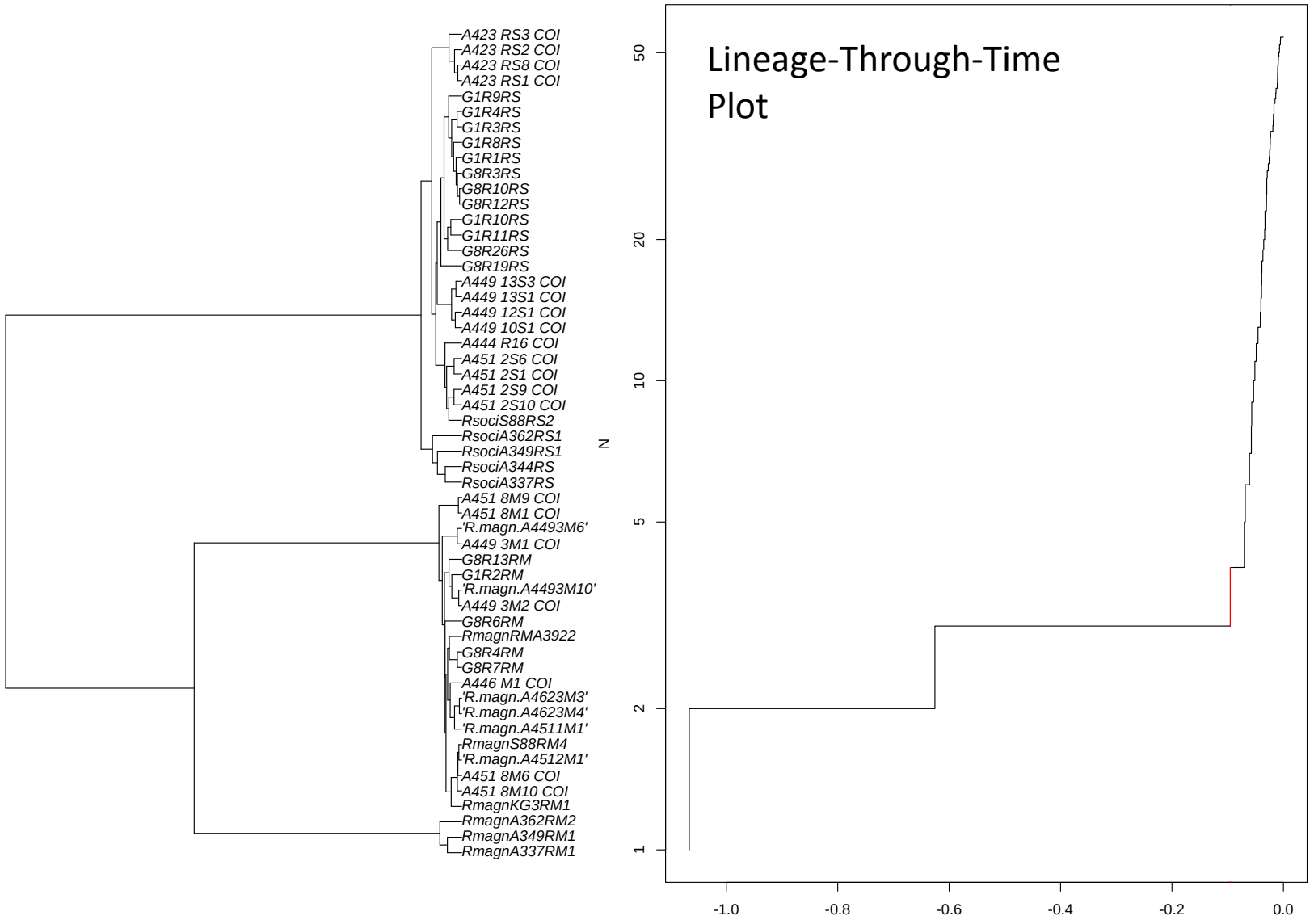
(Fujisawa & Barraclough, 2013: Systematic Biology)

C) Tree topology, Generalized Mixed Yule Coalescent model (GMYC)



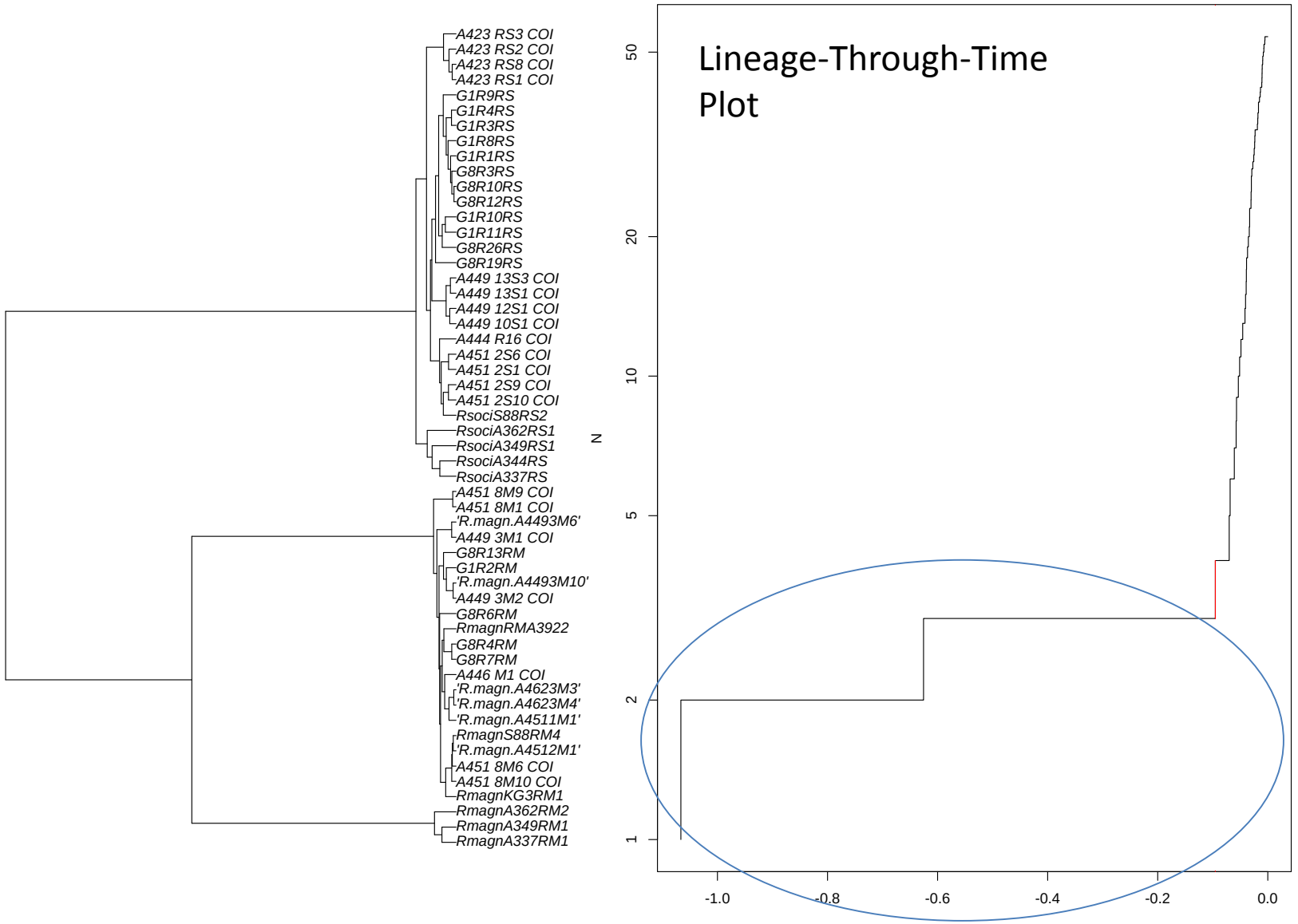
(Fujisawa & Barraclough, 2013: Systematic Biology)

C) Tree topology, Generalized Mixed Yule Coalescent model (GMYC)



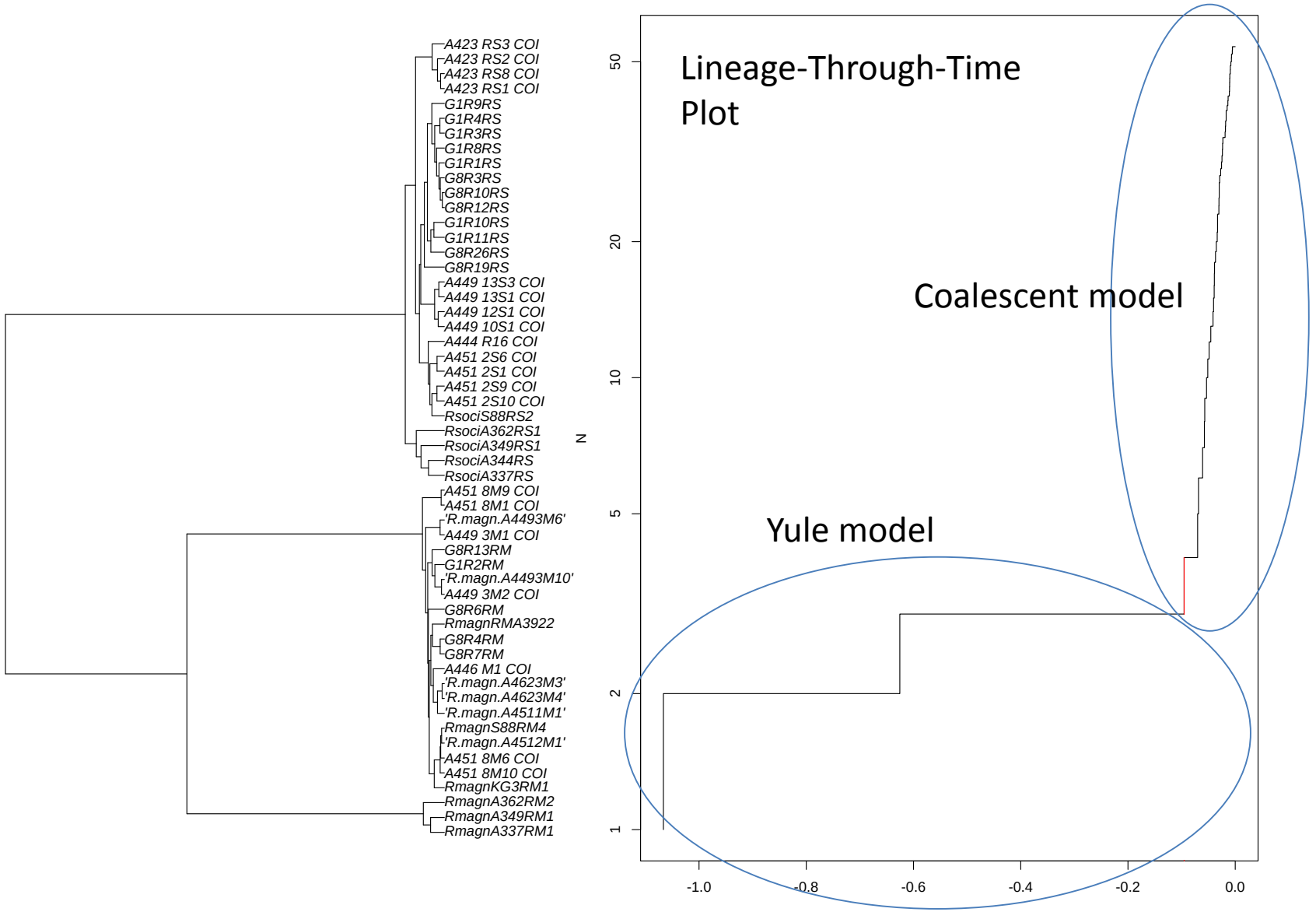
(Fujisawa & Barraclough, 2013: Systematic Biology)

C) Tree topology, Generalized Mixed Yule Coalescent model (GMYC)



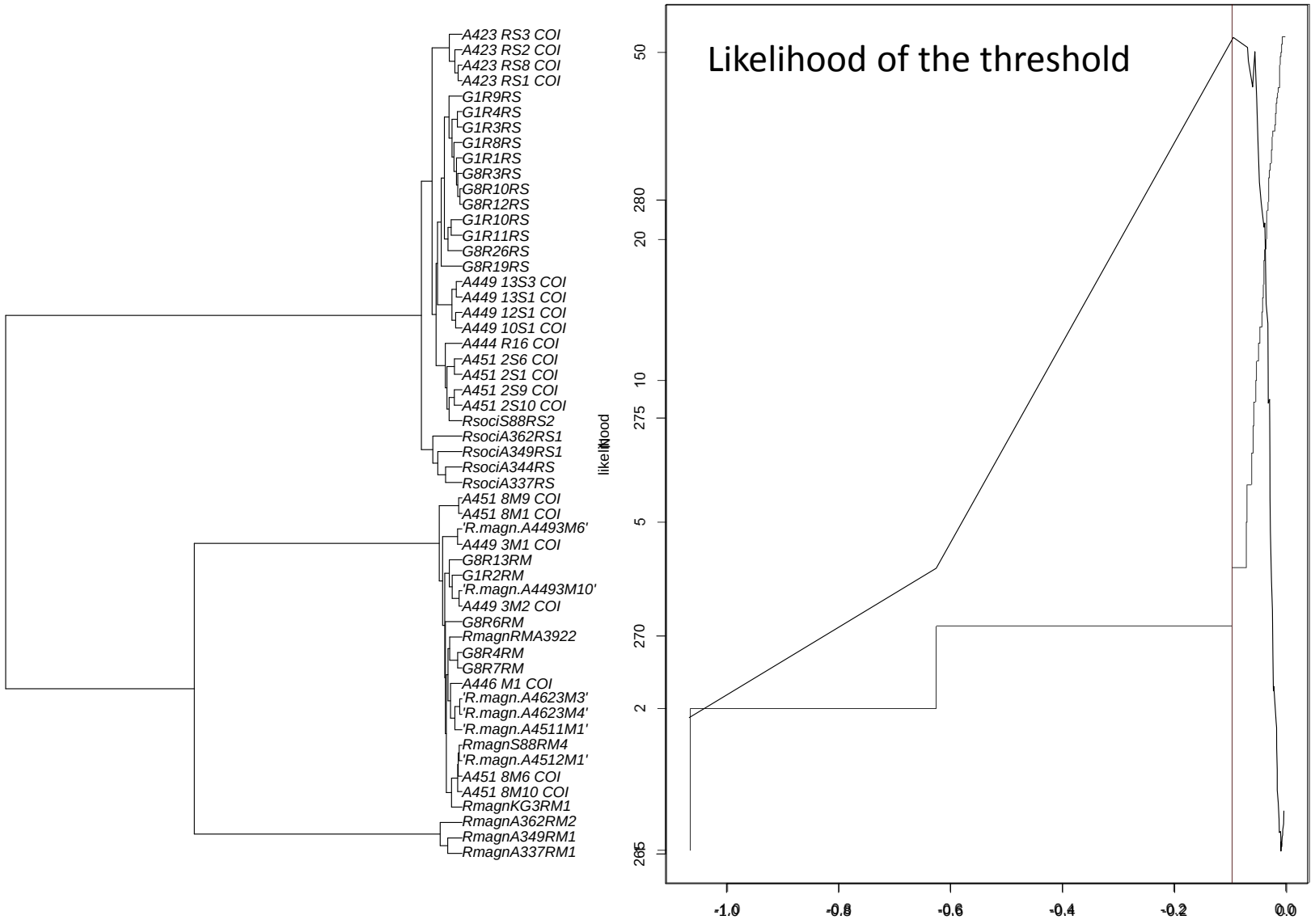
(Fujisawa & Barraclough, 2013: Systematic Biology)

C) Tree topology, Generalized Mixed Yule Coalescent model (GMYC)



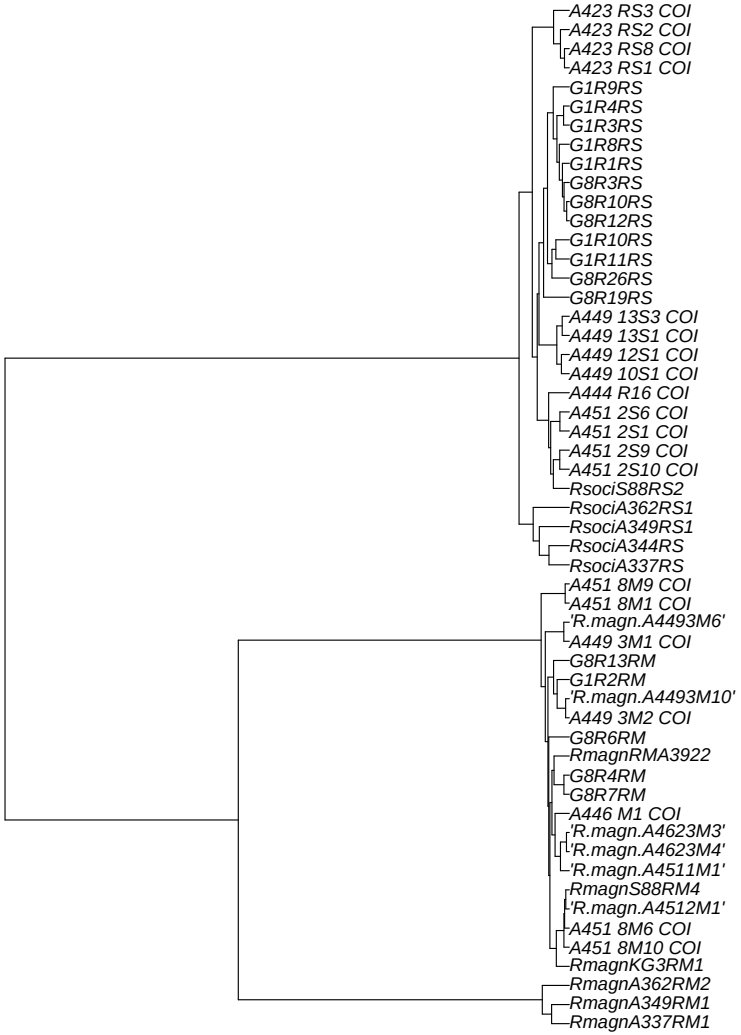
(Fujisawa & Barraclough, 2013: Systematic Biology)

C) Tree topology, Generalized Mixed Yule Coalescent model (GMYC)

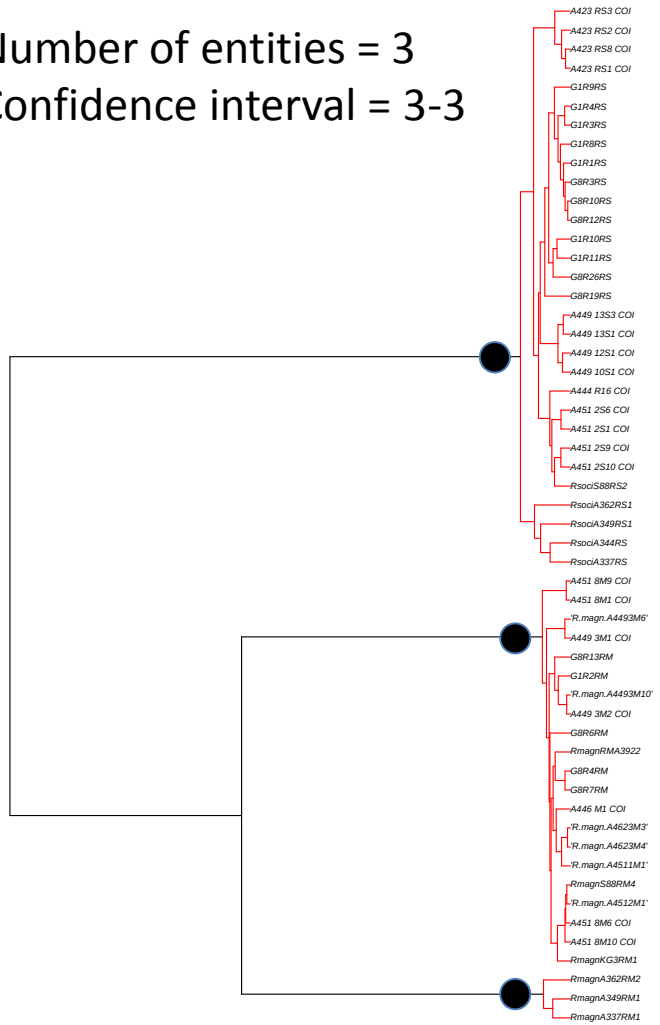


(Fujisawa & Barraclough, 2013: Systematic Biology)

C) Tree topology, Generalized Mixed Yule Coalescent model (GMYC)



Number of entities = 3
Confidence interval = 3-3

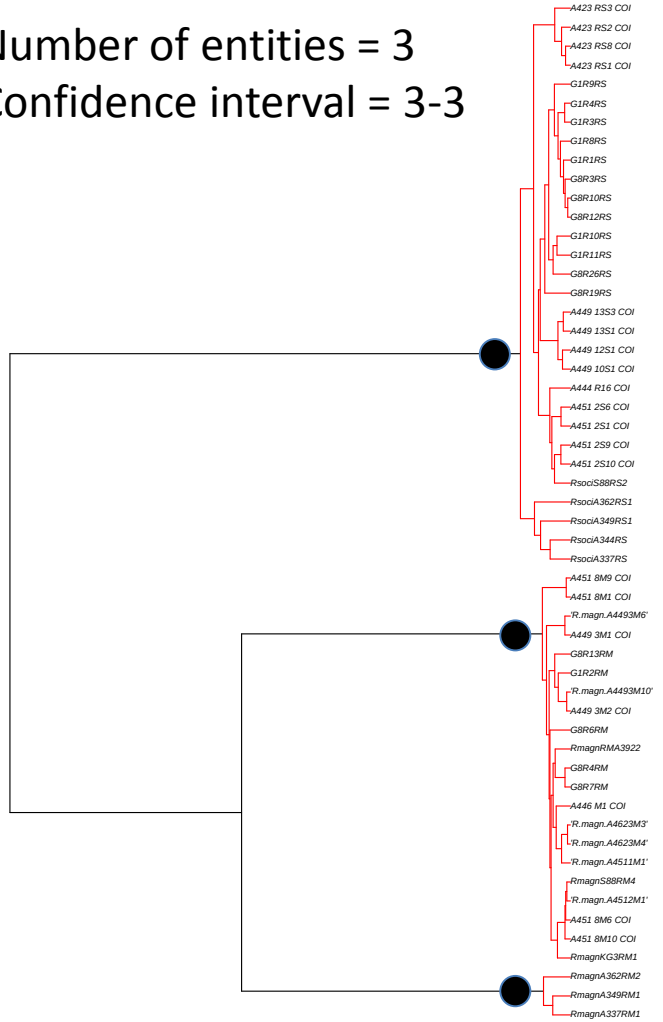


(Fujisawa & Barraclough, 2013: Systematic Biology)

C) Tree topology, Generalized Mixed Yule Coalescent model (GMYC)

package *splits*

Number of entities = 3
Confidence interval = 3-3



(Fujisawa & Barraclough, 2013: Systematic Biology)

Outline

1. Species lists: surveys & biomonitoring

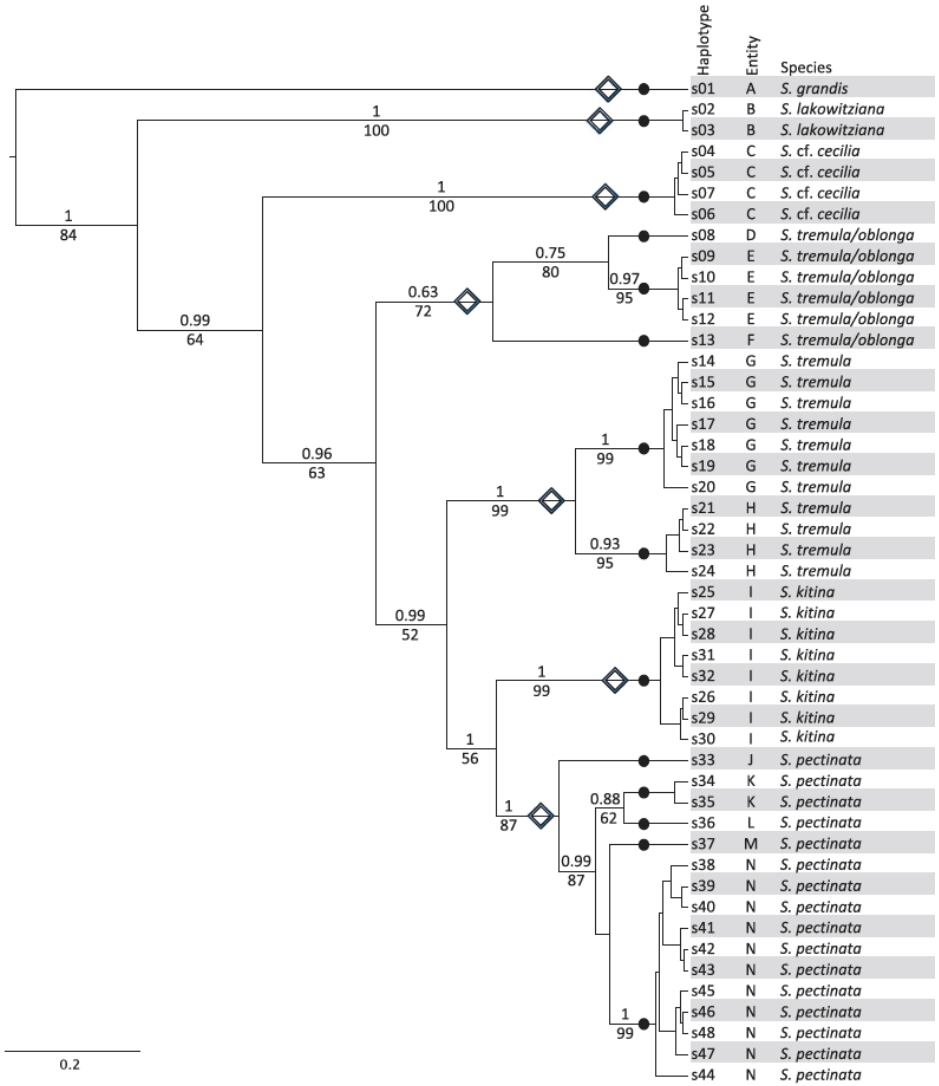
2. Species identification
   - Morphological taxonomy
   - DNA taxonomy
   - NGS and metagenetics

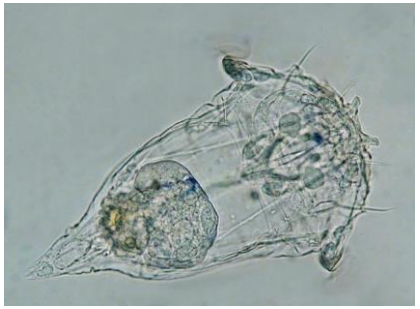
3. Inference

Is it useful?

2. Species identification: DNA taxonomy

The genus *Synchaeta* in alpine lakes





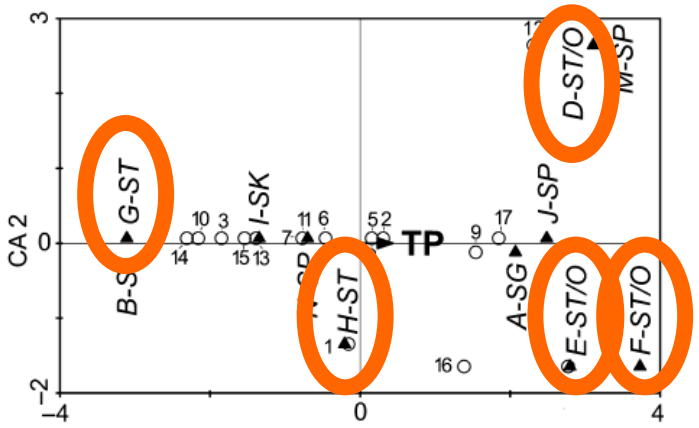
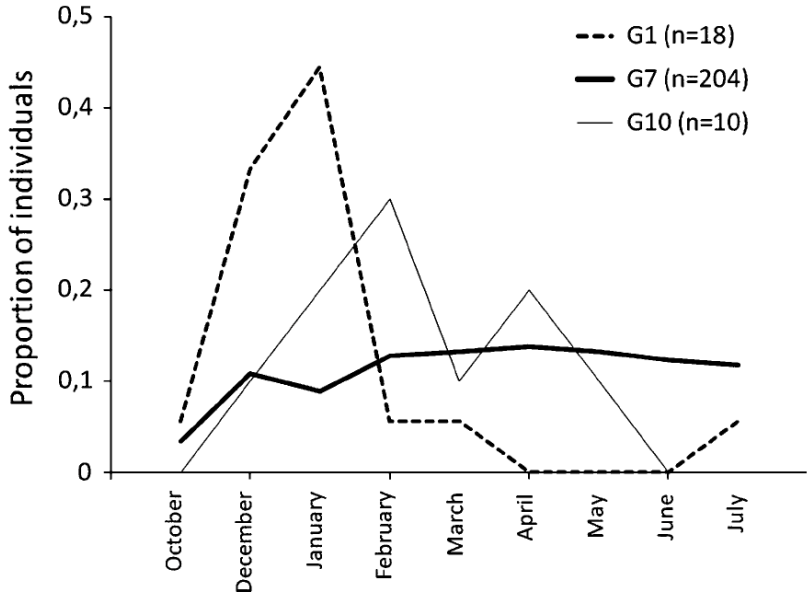


Fig. 3 CCA ordination ($\lambda_1 = 0.74$, $\lambda_2 = 1.00$) of *Synchaeta* entities identified by the generalised mixed Yule coalescent (GMYC) model. Open circles are waterbodies numbered according to Table 1, black triangles are GMYC entities named as in Fig. 2, the arrow indicates the environmental variable; *S. pectinata* = *SP*, *S. tremula* = *ST*, *S. tremula/oblonga* = *ST/O*, *S. kitina* = *SK*, *S. lakowitziana* = *SL*.

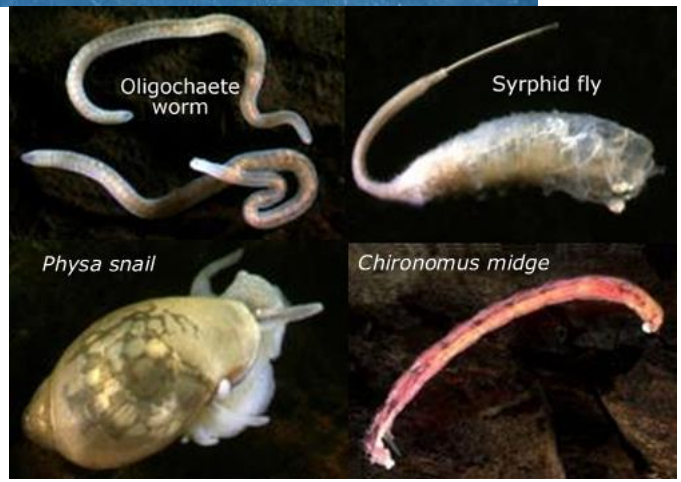
(Obertegger et al., 2012: Freshwater Biology)

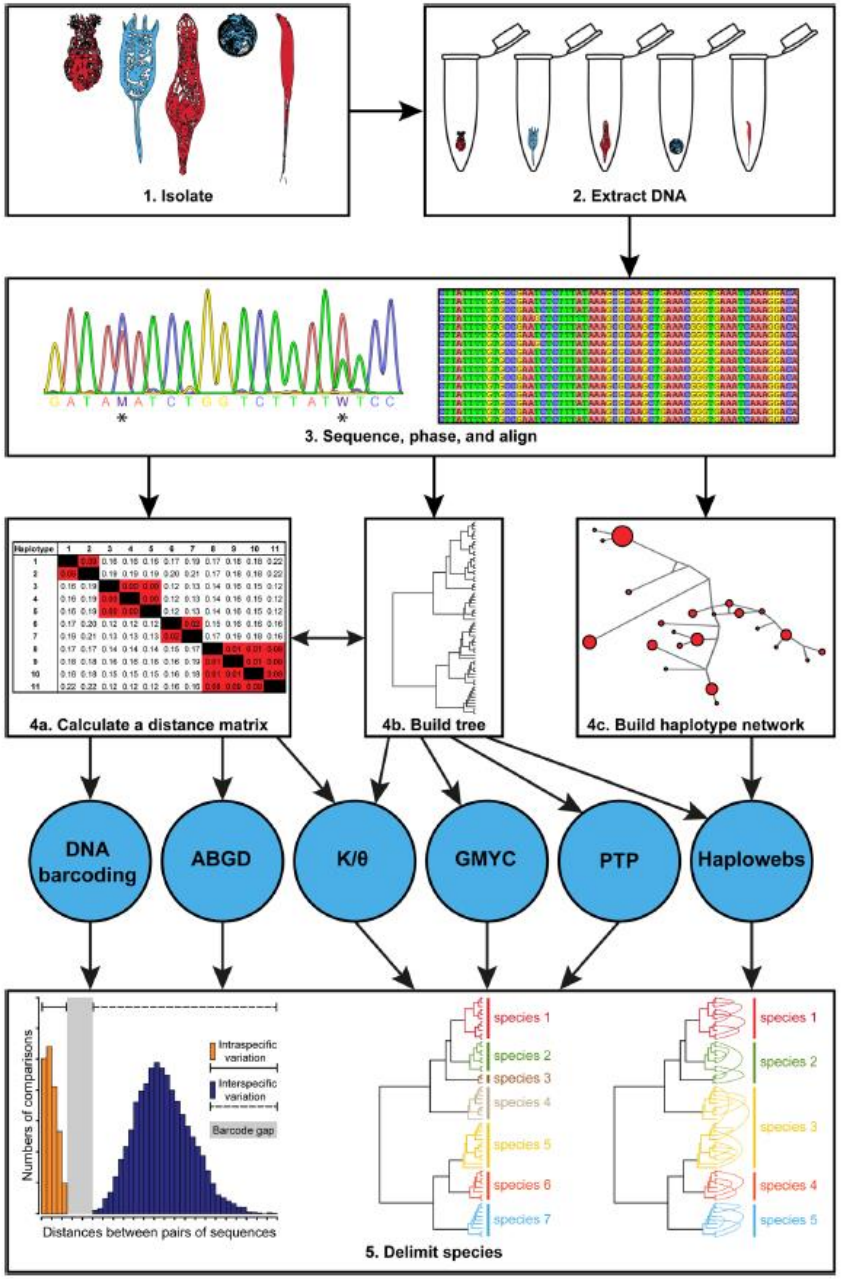


Temporal frequency of three co-occurring species of the *Baetis rhodani* complex

(Lucentini et al., 2011: Hydrobiologia)

2. Species identification: DNA taxonomy





Mar Biodiv
DOI 10.1007/s12526-015-0319-7

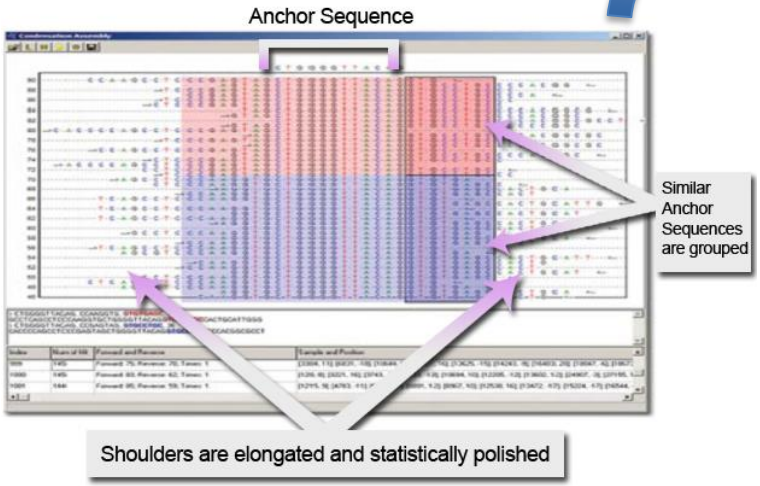
MEIOSCOOL

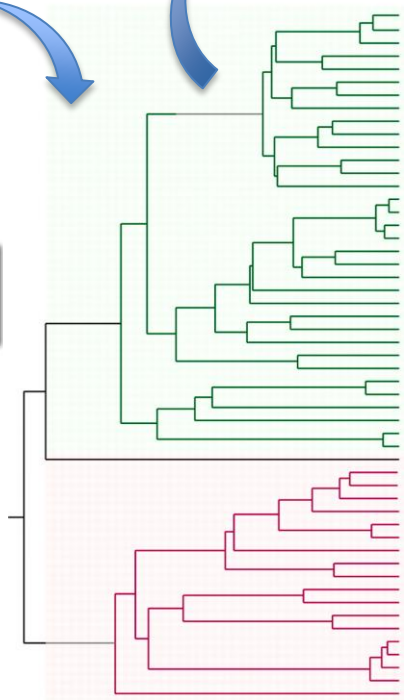
Guidelines for DNA taxonomy, with a focus on the meiofauna

Diego Fontaneto · Jean-François Flot · Cuong Q. Tang

2. Species identification: metagenetics

Metagenetics
Next Generation Sequencing
High Throughput Sequencing
Metabarcoding





Fixed threshold for OTUs
97%
99%

BLAST

2. Species identification: metagenetics

PNAS

The widely used small subunit 18S rDNA molecule greatly underestimates true diversity in biodiversity surveys of the meiofauna

Cuong Q. Tang[a,1], Francesca Leasi[a,b], Ulrike Obertegger[c], Alexander Kieneke[d], Timothy G. Barraclough[a], and Diego Fontaneto[a,e]

[a]Department of Life Sciences, Imperial College London, Ascot, Berkshire SL5 7PY, United Kingdom; [b]Department of Biodiversity, Earth and Environmental Science, Academy of Natural Sciences of Drexel University, Philadelphia, PA 19103; [c]Sustainable Agro-Ecosystems and Bioresources Department, Edmund Mach Foundation, Agricultural Institute at San Michele all'Adige Research and Innovation Centre, 38010 San Michele all'Adige, Italy; [d]Senckenberg am Meer Wilhelmshaven, German Centre for Marine Biodiversity Research, 26382 Wilhelmshaven, Germany; and [e]National Research Council, Institute of Ecosystem Study, 28922 Verbania Pallanza, Italy

BMC Ecology

METHODOLOGY ARTICLE Open Access

Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol

Mehrdad Hajibabaei*, Jennifer L Spall, Shadi Shokralla and Steven van Konynenburg

MOLECULAR ECOLOGY

Molecular Ecology (2012) 21, 2039–2044

NEWS AND VIEWS

OPINION

Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing

DONALD J. BAIRD* and MEHRDAD HAJIBABAEI†

nature COMMUNICATIONS

ARTICLE

Received 15 Nov 2012 | Accepted 12 Apr 2013 | Published 21 May 2013

DOI: 10.1038/ncomms2881

Whole-community DNA barcoding reveals a spatio-temporal continuum of biodiversity at species and genetic levels

Andrés Baselga[1], Tomochika Fujisawa[2,3,†], Alexandra Crampton-Platt[2,3], Johannes Bergsten[2,3,†], Peter G. Foster[2], Michael T. Monaghan[2,3,4] & Alfried P. Vogler[2,3]

Outline

1. Species lists: surveys & biomonitoring

2. Species identification
   - Morphological taxonomy
   - DNA taxonomy
   - NGS and metagenetics

3. Inference

3. Are results comparable?

Taxon	number of species
Acoela	20
Nemertodermatida	6
Platyhelminthes	65
Gnathostomulida	3
Rotifera	32
Cycliophora	1
Polychaeta	6
Nematoda	95
Priapulida	2
Kinorhyncha	6
Gastrotricha	32
Tardigrada	6
Copepoda	165

(Curini-Galletti et al., 2012: PLoS ONE)

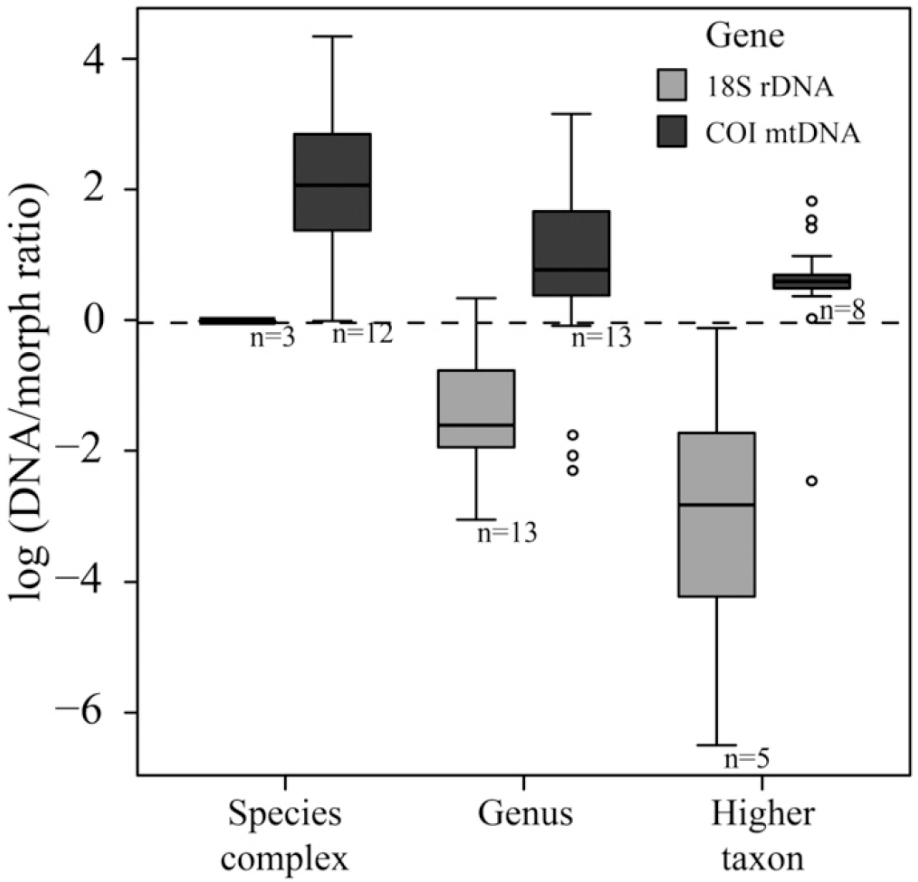
3. Are results comparable?

Taxon	number of species	OTUs from NGS of 18S fragment
Acoela	20	20
Nemertodermatida	6	10
Platyhelminthes	65	1,000
Gnathostomulida	3	30
Rotifera	32	2
Cycliophora	1	0
Polychaeta	6	500
Nematoda	95	5,000
Priapulida	2	2
Kinorhyncha	6	100
Gastrotricha	32	30
Tardigrada	6	50
Copepoda	165	500

(Curini-Galletti et al., 2012: PLoS ONE)

(unpublished)

3. Are results comparable?



Meiofauna

12,000 individuals sequenced

55 taxa

8 phyla

COI vs 18S

Fixed threshold
ABGD
K/theta
GMYC

(Tang et al., 2012: PNAS)

Perspectives on future work

(1) Create a reference database for DNA barcoding

DNA sequences from all relevant organisms in Italian lakes

(2) Compare results from MetaBarcoding and current WFD approaches

in a selection of well-known Italian lakes

(3) Produce a workflow to integrate MetaBarcoding into WFD

to be used for rapid and efficient water quality assessment

Advantages

(1) Cheaper and less time consuming

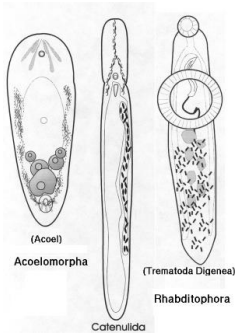
DNA sequences cost less than the time of taxonomists

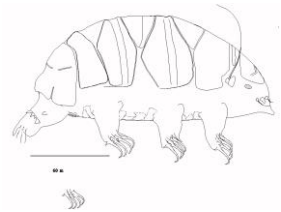
(2) Quantitative DNA data available forever

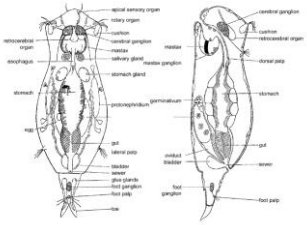
DNA sequences are freely available and can be used for other analyses

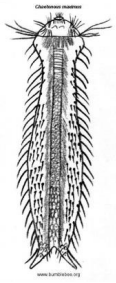
(3) Target the work of taxonomists

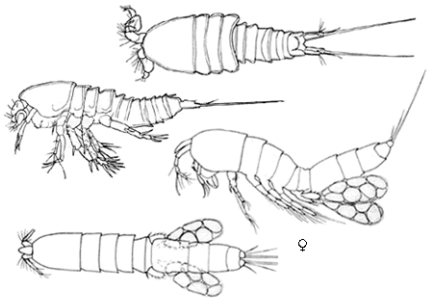
concentrate efforts only in samples with previously unknown DNA sequences





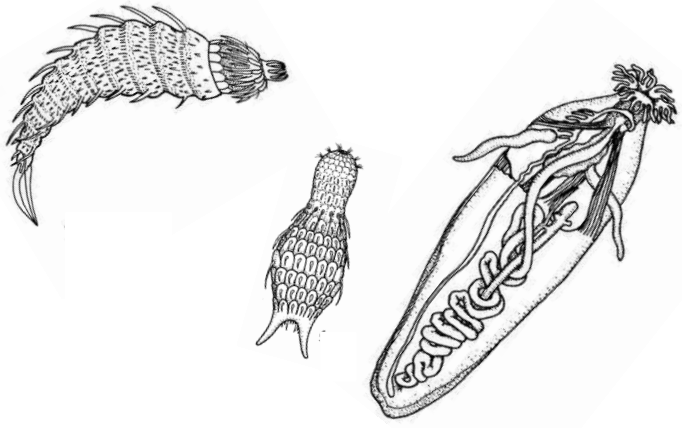


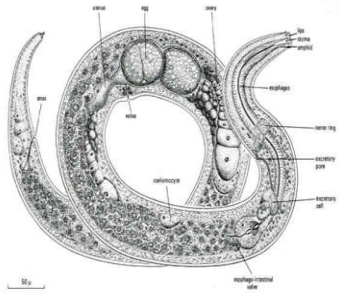


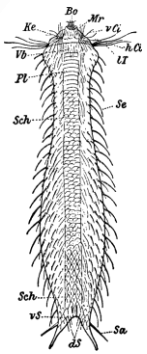


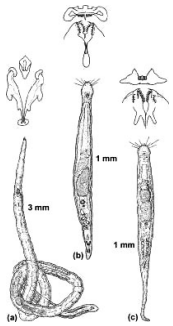
After: Huys & Boxshall, 1991

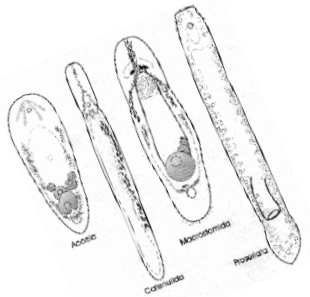
Tassonomia molecolare per razionalizzare il monitoraggio biologico











Diego Fontaneto
CNR-ISE, Verbania Pallanza, Italy
d.fontaneto@ise.cnr.it